5.3 Estimates of Dose-Response Rates for Acute and Chronic Illness

To place any credence in the estimates presented in this section, one must believe that stochastic factors play a role in dose-response functions. Stochastic disturbances may have a greater or a lesser part to play than systematic biological, physical, economic, or social influences, but they nevertheless have a part. If all influences were entirely deterministic, the statistical procedures employed here (as well as all of epidemiology) would be unnecessary and redundant: all one would have to do to ascertain the values of the influences is go to the laboratory and perform the relevant measures. In fact, single observations on the phenomena of interest would suffice: if the observations conformed to the proposition, one would accept the proposition for now. Otherwise, it would be rejected, Biomedical research employs both laboratory and human population studies (and several different variants within each of these general classifications) to come to grips, most often with less then iron firmness, with dose-response functions. The use of these approaches and their variants is an admission that the functions involve significant stochastic elements.

Reference is made to rates rather than functions in the subtitle of this section because the empirical results reported apply only to changes in measured illness for one-unit changes in the explanatory variables of interest at the mean values of these dependent and explanatory variables. These changes could properly describe an entire dose-response function if and only if that function were linear in the original variables. Throughout the estimation procedure, we have employed linear functions for an assortment of reasons, not the least of which is that there appears to be no strong analytical or empirical precedence for doing otherwise with the generalized measures of ill-health we are using. We don't know whether the air pollution dose-health response function is supposed to be increasing at an increasing or a decreasing rate over a given interval. A linear function is the best available compromise between these two possibilities. The linear form is easily interpreted at a glance and, furthermore, relative to other readily estimated forms such as the multiplicative, it does not attenuate the potential influence of observations having extreme values. In the absence of knowledge about the functional form of the relationship one is estimating, the use of multiplicative and similar forms effectively reduces the variation of the sample and thus will often allow one to explain a larger proportion of the variation in the (rescaled sample. For purposes of the present study, since we lack prior knowledge of the form of the dose-response functions, we wish to provide the extremes of good and ill health, and pristine and filthy air, full rein. This reluctance to reduce the influence of outliers, when combined with our use of data on individual human being rather than group averages, means that we reduce, if not completely deny, our chances of explaining large proportions of the variation among our basic observational units in acute and chronic illness.

Tables 5.6a and 5.6b present estimates for household heads of dose-response relations for acute illness and Tables 5.7a and 5.7b do the same for chronic illness. So as to reduce the extent to which cumulative exposures to outdoor air pollutants are unaccounted for, all the estimated

chronic illness expressions employ as basic units of observation only household heads who have always resided in one state. This restriction is imposed for all chronic illness estimates throughout the chapter.

Substantial care has been taken to assure that all explanatory variables have either always been outside the household head's domain of control or have been established by his actions prior to the period being considered. Thus, variables such as the head's age, where he grew up, his father's education, and past financial status, his sex and race, and the cold, air pollution, and the ultraviolet radiation to which he is exposed at a particular location are matters over which he never has and never will exercise anything but the most trivial influence. They are inalterable. Other variables such as the severity of any disabilities he has, and his education, marital status, and family size were certainly influenced by his decisions. However, the impact of past decisions on the current values of these variables will, for nearly all adults, overwhelm any potential impact of decisions made within any current 12 month period. The economic sector within which one is employed and the rooms per family members in one's housing are perhaps subject to more immediate control but, for the great bulk of people, are not very quickly or readily adjusted. Assertions of predetermination are clearly inaccurate for most of the life-style variables. One's current cigarette consumption, exercise, and dietary habits, etc., are quickly adapted to changing circumstances. Yet one might also reasonably argue that even these current adaptations are isomorphic to acquired habits, and can thus be employed as proxies for these predilections. In fact, for items such as medical insurance, food and cigarettes, there is abundant evidence in the empirical consumer demand literature that the quantities individuals consume are quite insensitive to price changes, at least for the range of price changes likely to occur in a year. Similarly, these habits tend to persist for some time in the face of substantial yearly income changes. Finally, introspection says that one's religious and risk aversion attitudes are the result of the accumulated experiences and learning of a lifetime rather than a momentary diversion that will serve only until a new fad comes to one's attention.

A rather large data set like the SRC survey, when joined with a quite sparse set of a priori propositions with which to restrict the expressions to be estimated, leads one into temptation. In particular, using an unchanging set of sample observations, one is tempted to add and delete variables and try assorted functional forms until a result is obtained that, on statistical grounds alone, looks good; that is, the coefficients attached to the explanatory variables all have common sense or a priori acceptable signs and are generally statistically significant at high levels. Moreover, summary statistics such as the coefficient of determination are high and standard errors of estimate are low. Quite frequently, the results of this "data-grubbing" are reported without any description of the manipulations lying behind them. As is well-known, this practice can introduce substantial biases into estimated coefficients. In the words of Selvin and Stuart (1966, p. 21):

> "... any preliminary search of data for a model, even when the alternatives are predesigned, affects the probability levels of

> all subsequent tests based on that model on the same data, and in no very simple way, and also affects the characteristics of subsequent estimation procedures. The only valid course is to use different data for testing the model dredged from the first set of data."

We have not conformed absolutely to this dictum, but have nevertheless followed it rather closely.[10/]

In Tables 5.6a, 5.6b, 5.7b, each estimated expression is numbered, with each number in each table corresponding to an entirely new sample drawn at random from the entire SRC population sample or that portion of the SRC sample meeting certain imposed conditions. Thus, for example, in Table 5.6a expressions (1A), (1B), and (1C), are estimated from the same set of observations but the expressions (1) and the expressions (2) are estimated from entirely different samples. Since the availability of variables in the SRC data set can differ greatly from year-to-year, and the definitions of variables can differ slightly, it is not possible to exploit formal statistical tests for replication. Nevertheless, if the different samples do yield similar results for a particular set of variables, a dimension is added to the estimation procedure that undeniably adds information and confidence in the results.

Even though a modicum of something resembling data-grubbing is present in the estimation of expressions like (1A), (1B), and (1C) in Table 5.6a, it does not involve anything more than using the same data set to reestimate expressions in which nothing other than the air pollution variables has been changed. Thus, though (1) in Table 5.6a involves three expressions, only three "runs," with one run for each combination of air pollution variables, was performed.

Table 5.5 is a table of simple correlation coefficients for a representative sample. These coefficients, of course, differed from one sample to another, but the table provides a good idea of the general patterns of intercorrelation among the variables that were estimated by the various samples. As a glance at the table shows, there is very little linear association between the air pollution variables and any single other variable used to explain acute and chronic illness. No one of these other explanatory variables linearly accounts for more than 23 percent of the variance of an air pollution variable, and, in most cases, the variance accounted for is considerably below ten percent. Similarly, the intercorrelation among variables other than the air pollution variables tends to be very low. This, of course, does not mean that strong nonlinear associations between single variables are absent. Neither does it mean that close associations between the air pollution variables and linear or nonlinear combinations of other explanatory variables are not present. Although there exist some statistics that purport to test for these latter two possibilities, we have not employed them in this report. We thus proceed as if the fact that linear associations between single explanatory variables are typically low implies that multicollinearities among all explanatory variables (except for the air pollution variables) are unlikely to inflate the standard errors of coefficients, thereby causing certain

Table 5.5

Matrix of Simple Correlation Coefficients for a 1971 Representative Dose-Response Function Sample

	LDSA	EDUC	CIGE	EXER	FOOD	RISK	AGEH	DSAB	FMSZ	SEXH	INSR	SVGS	CHEM	CITY
LDSA	1.000	-0.139	-0.112	-0.075	-0.202	-0.132	0.201	0.700	0.037	-0.117	-0.301	-0.175	0.029	-0.008
EDUC		1.000	0.001	0.230	0.442	0.454	-0.155	-0.153	-0.191	0.217	0.401	0.354	0.201	0.057
CIGE				0.006	0.077	-0.268	-0.091	-0.067	0.126	0.303	0.186	0.018	-0.037	0.303
EXER					0.054	0.134	-0.106	-0.059	0.038	0.167	0.151	-0.150	-0.024	-0.025
FOOD						0.454	0.157	-0.172	-0.369	0.311	0.414	0.501	0.059	0.012
RISK							-0.021	-0.200	0.173	0.427	0.678	0.043	-0.032	0.068
AGEH								0.231	-0.081	0.018	0.001	0.284	0.076	-0.094
DSAB									-0.035	-0.184	-0.325	-0.165	-0.024	-0.048
FMSZ										0.081	-0.206	-0.188	-0.004	0.047
SEXH											0.217	0.215	0.033	0.088
INSR												0.418	0.033	-0.027
SVGS													0.065	-0.058
CHEM														0.035
POOR	0.156	-0.137	-0.124	0.024	-0.239	-0.133	0.115	0.137	0.108	0.030	-0.060	-0.165	-0.052	0.076
FEDU	-0.164	0.188	0.137	0.169	0.501	0.214	-0.176	-0.135	-0.182	0.147	0.177	0.119	-0.012	0.100
RACE	-0.160	0.419	0.306	0.170	0.732	0.447	0.180	-0.136	-0.235	0.353	0.413	0.414	0.050	-0.017
ACUT	0.006	-0.013	0.023	-0.080	0.289	-0.053	0.026	0.155	-0.059	-0.071	0.046	-0.024	-0.018	-0.055
TSPT	0.096	-0.012	-0.084	-0.084	0.044	0.042	-0.085	-0.043	0.034	0.110	0.066	0.155	0.091	0.042
TSPN	0.153	-0.058	0.079	-0.106	0.050	0.136	-0.097	0.080	0.055	0.091	0.089	0.231	0.100	0.074
TSPM	0.124	-0.024	0.079	-0.119	0.016	0.066	-0.086	0.150	0.040	0.133	0.047	0.172	0.088	0.056
SULT	-0.002	0.053	0.057	-0.126	0.018	0.069	-0.031	0.066	0.021	0.156	-0.054	0.108	0.015	0.072
SULN	-0.006	0.113	0.087	-0.129	0.072	0.136	-0.085	0.043	0.025	0.097	0.042	0.217	0.073	0.068
SULM	0.002	-0.021	0.045	-0.142	-0.009	0.058	-0.044	0.045	0.013	0.169	-0.031	0.119	0.031	0.071

	FEDU	RACE	ACUT	TSPT	TSPN	TSPM	SULT	SULN	SULM
POOR	-0.285	-0.130	-0.094	0.054	0.067	0.056	0.004	0.038	0.005
FEDU		0.326	0.030	0.094	0.039	0.075	0.039	0.052	0.065
RACE				-0.071	0.008	0.056	-0.012	-0.082	-0.100
ACUT				0.155	0.074	0.096	0.119	0.122	0.122
TSPT					0.922	0.970	0.441	0.742	0.658
TSPN						0.976	0.652	0.897	0.821
TSPM							0.622	0.861	0.800
SULT								0.868	0.945
SULN									0.939

Table 5.6a

Dose-Response Rates for ACUT: Unpartitioned Samples

Year	(1A)		(1B)		(1C)		(2)		(3)	
Variable	β 1967	s	β 1967	s	β 1967	s	β 1968	s	β 1969	s
DSAB	20.541	5.862	21.140*	5.947	21.520*	5.854				
LDSA							47.04*	16.08	3.252	12.290
AGEH	-2.486	1.650	-2.068	1.246	-1.895	1.637	-1.306	1.456	-1.208	1.097
EDUC	4.086	13.344	4.155	13.540	4.462	13.300				
MARR	-12.561	81.952	-8.362	81.660	-21.280	80.500	16.610	35.560	0.065	29.11
POOR	-24.264	40.419	-24.060	40.800	-26.120	40.180	-29.80	34.03	-52.320*	27.030
RACE	-87.746*	46.328	-95.090*	49.95	-109.900*	53.220				
SEXH	-17.564	87.082	-7.666	86.450	-20.370	85.480				
EXER									-66.732*	34.930
FOOD	-0.062*	0.039	-0.063*	0.033	-0.066*	0.037	-0.056*	0.023	-0.071	0.175
NCIG	17.943	11.801	18.520	12.010	20.170*	11.760	16.130*	9.844		
RELG	-12.561	81.958								
RISK	-9.392	16.670	-11.170	17.190	-13.770	16.720	-17.676*	12.330	-25.960*	9.668
INSR	20.84	59.05	15.150	59.31	13.380	60.380	88.710**	47.090	67.510**	37.420
CHEM										
DENS									1.127	7.429
NOXT										
NOXM										
NOXN										
SULT	1.857*	1.033								
SULM			1.488*	0.733			1.518*	0.925	-1.199	0.951
SULN					0.722*	0.372				
TSPT	-0.432	0.681								
TSPM			-0.442	0.648			-0.963	0.606	1.122*	0.765
TSPN					-0.120	0.261				

(continued)

Table 5.6a
(continued)

Year	(1A)		(1B)		(1C)		(2)		(3)	
Variable	β	1967 s	β	1967 s	β	1967 s	β	1968 s	β	1969 s
Constant	410.960		322.546		320.309		447.874		283.201	
R^2	0.307		0.296		0.313		0.175		0.182	
S.E.	164.745		166.030		164.108		317.210		264.023	
F	(13,80) = 4.731		(13,80) = 4.594		(13,80) = 4.800		(10,389) = 6.139		(10,389) = 5.473	
η_{NOX}										
η_{SUL}	0.308		0.353		0.544		0.326			
η_{TSP}									0.474	

(continued)

Table 5.6a
(continued)

Year	(4)		(5)		(6)		(7)	
Variable	β 1970	s	β 1971	s	β 1972	s	β 1973	s
DSAB					180.800*	26.550		
LDSA	76.490	13.920	47.990*	14.590			19.340	14.740
AGEH	0.485	1.153	2.542*	1.199	1.411	1.563	0.355	1.051
EDUC			-15.800*	8.370			2.550	7.818
POOR	-63.200*	27.360	-49.260*	26.640	128.100*	41.550	-14.190	28.110
RACE					4.435	22.680	-36.450*	20.460
SEXH	18.490	31.070	-85.170*	31.090	20.740	43.290	-123.600*	29.390
CIGE								
EXER	-48.620	30.260	-30.150	31.560				
FOOD			-0.030	0.021				
NCIG								
RELG								
RISK	-37.120*	9.872	-21.796*	9.876				
INSR	9.885	36.780	7.439	40.330				
CHEM	-3.280	55.020			-99.730	110.300	54.410	76.740
DENS					3.529*	1.597	0.056	1.075
NOXT								
NOXM								
NOXN					-0.972	0.684	0.223*	0.124
SULT								
SULM	2.520*	1.104	2.257	2.259			-0.361	3.305
SULN								
TSPT								
TSPM			1.453*	0.764	1.782*	0.780	-0.249	0.314
TSPN								

(continued)

Table 5.6a
(continued)

Year	(4)	(5)	(6)	(7)
Variable	β 1970 s	β 1971 s	β 1972 s	β 1973 s
Constant	305.260	172.464	-78.317	175.040
R^2	0.123	0.123	0.169	0.095
S.E.	262.333	252.936	394.533	254.413
F	(9,390) = 6.104	(11,388) = 4.926	(9,390) = 8.836	(11,388) = 4.435
η_{NOX}				0.618
η_{SUL}	0.361	0.518		
η_{TSP}			0.497	

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

Table 5.6b

Dose-Response Rates for ACUT: Partitioned Samples

Year	(1) 1967		(2) 1969		(3) 1969		(4) 1970	
	Always lived in 1 state		*RINC = ≤ $7,500*		*3 ≤ NCIG ≤ 6*		*1 ≤ DSAB ≤ 3*	
Variable	β	s	β	s	β	s	β	s
DSAB	42.056*	6.538						
LDSA			111.200*	37.590	17.960	13.810	-94.990*	34.430
AGEH	0.384*	0.187	2.889	3.488	-2.383	1.435	1.215	
EDUC	2.716	2.302						
MARR	-17.037	85.185	117.800	75.310	35.290	38.710	56.630	99.450
POOR	31.832	46.812	-116.400	78.780	5.323	31.210	-66.900	86.310
RACE	-60.549	53.583						
SEXH	13.327	9.005						
FOOD	-0.061	0.057	-1.648*	0.559	-0.084	0.218	-0.168	0.519
NCIG	5.643	3.239*	36.940*	22.710	34.030*	18.620		
RISK	-4.047	17.600	32.950	27.840	-4.700	12.380	1.938	35.750
INSR	-75.286	70.361	80.820	86.440	-71.390*	42.490	-54.560	117.000
SULM	-0.992	7.631	5.135*	3.020	0.007	0.831	0.114	2.930
TSPM	1.765*	0.865	-4.031	3.020	-0.594	0.480	1.215	2.210
EXER							200.600*	125.600
DENS							-31.710	21.700
Constant	121.290		566.723		165.600		482.897	
R^2	0.152		0.186		0.076		0.122	
S.E.	352.420		443.738		243.090		449.633	
F	(14,306) = 6.621		(10,150) = 3.431		(10,268) = 2.191		(10,114) = 1.585	
NSUL			0.565					
NTSP	0.952							

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

Table 5.7a

Dose-Response Rates for LDSA: Unpartitioned Samples[a]

Year	(1)		(2)		(3a)		(3b)		(5A)	
Variable	β 1967	s	β 1968	s	β 1970	s	β 1970	s	β 1971	s
DSAB			3.286*	0.227	0.554**	0.035	0.550**	0.035	0.808**	0.049
AGEH	0.003	0.007	-0.002	0.007	0.005	0.004	0.005	0.004	0.007*	0.004
CITY										
EDUC	0.079	0.054	0.170	0.416	0.013	0.029	0.001	0.029	-0.057	0.030
FEDU					-0.044	0.037	-0.043	0.037	-0.044	0.035
MARR	0.204	0.284								
POOR	0.188	0.157	0.135	0.163	-0.069	0.103	-0.065	0.103	0.086	0.096
RACE	0.344	0.200			0.072	0.488	0.088	0.487	-0.057	0.119
SEXH	0.410	0.297			0.139	0.114	0.132	0.113	0.233**	0.109
FOOD	-0.7×10^{-4}	0.26×10^{-3}	0.002*	0.001	-0.902	0.975	-0.924	0.973	-0.13×10^{-3}	0.81×10^{-4}
NCIG	0.023	0.047	-0.089*	0.041						
RISK	-0.009	0.006								
INSR	-0.152	0.245	-0.336*	0.218	-0.454*	0.129	-0.459*	0.129	-0.496*	0.125
CHEM					-1.645**	0.575	-0.097	0.575	-0.002	0.916
NOXT										
NOXM										
NOXN										
SULM	0.0036	0.0025	0.0067*	0.0035						
SULN										
TSPM	0.0021	0.0037	-0.0036	0.0024	0.0028*	0.0011	0.0018*	0.66×10^{-3}	0.0019	0.0017
TSPN										
Constant	-0.636		0.631		2.980		2.924		0.265	
R^2	0.094		0.371		0.525		0.526		0.530	
S.E.	0.835		0.736		0.964		0.963		0.904	
F	(12,134) = 2.158		(9,390) = 25.580		(11,388) = 38.920		(11,388) = 39.170		(11,388) = 39.800	
η_{NOX}										
η_{SUL}					0.278		0.341		0.268	
η_{TSP}										

(continued)

Table 5.7a

(continued)

Year	(5B)		(6A)		(6B)	
Variable	β 1971	s	β 1972	s	β 1972	s
DSAB	0.809**	0.049				
AGEH	0.007*	0.004	0.023*	0.005	0.020*	0.005
CITY			-0.057	0.050	-0.085	0.045
EDUC	-0.058	0.030	-0.081	0.045	-0.125**	0.043
FEDU	-0.043	0.035	0.007	0.050	0.048	0.055
MARR						
POOR	0.088	0.097	0.104	0.147	0.116	0.145
RACE	-0.054	0.119	-0.272	0.160	-0.220	0.154
SEXH	0.240**	0.109	-0.156	0.141	-0.182	0.142
FOOD	-0.13×10^{-3}	0.81×10^{-4}				
NCIG						
RISK						
INSR	-0.499*	0.125				
CHEM	-0.016	0.917				
NOXT						
NOXM			-0.0007	0.0017		
NOXN					0.14×10^{-3}	0.89×10^{-3}
SULM						
SULN						
TSPM			0.0028	0.0027		
TSPN	0.59×10^{-3}	0.73×10^{-3}			0.0030*	0.0013
Constant	0.181		0.701		1.054	
R^2	0.529		0.119		0.134	
S.E.	0.905		1.347		1.315	
F	(11,388) = 39.680		(9,390) = 5.879		(9,390) = 6.706	
η_{NOX}						
η_{SUL}						
η_{TSP}			0.376		0.630	

(continued)

Table 5.7a
(continued)

Year	(7A)		(7B)		(8A)		(8B)	
Variable	β	1973 s	β	1973 s	β	1974 s	β	1974 s
DSAB								
AGEH	0.017*	0.044	0.017*	0.004	0.017*	0.005	0.017*	0.005
CITY	0.155	0.140	0.180	0.140				
EDUC	-0.122*	0.043	-0.128*	0.043	-0.118**	0.049	-0.111**	0.049
FEDU	0.059	0.052	0.057	0.053				
MARR					-0.291	0.221	-0.287	0.221
POOR	0.050	0.143	0.029	0.142	0.302*	0.151	0.305*	0.151
RACE	-0.208	0.154	-0.202	0.155				
SEXH	-0.207	0.141	-0.209	0.141	-0.001	0.230	-0.001	0.230
FOOD					-0.060	0.066	-0.067	0.067
NCIG								
RELG					-0.459	0.284	-0.457	0.284
INSR								
CHEM					-0.161	0.327	-0.131	0.325
NOXT								
NOXM					0.0023	0.0017		
NOXN							0.0046*	0.0025
SULM	0.0033	0.0037			-0.0047	0.0062		
SULN			0.003	0.002			0.0002	0.0022
TSPM	0.0017	0.0015			0.0008	0.0028		
TSPN			0.0004	0.0017			-0.0007	0.0019
Constant	0.309		0.505		-0.687		-0.828	
R^2	0.106		0.109		0.118		0.112	
S.E.	1.303		1.303		0.966		0.964	
F	(9,390) = 5.785		(9,390) = 5.290		(11,214) = 4.591		(11,214) = 4.693	
η_{NOX}					0.363		1.143	
η_{SUL}								
η_{TSP}								

(continued)

Table 5.7a
(continued)

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

[a]All observations in this table are limited to individuals who have always lived in one state.

Table 5.7b

Dose-Response Rates for LDSA: Partitioned Samples[a]

Year	(1) 1971 50-cities		(2) 1969		(3A) 1972 AGEH ≥ 45		(3B) 1972 AGEH ≥ 45		(4A) 1972 AGEH ≥ 45 & MILE ≤ 15		(4B) 1972 AGEH ≤ 45 & MILE ≤ 15	
Variable	β	s	β	s	β	s	β	s	β	s	β	s
DSAB	-0.168**	0.068	2.462*	0.108					0.904**	0.054	0.897**	0.057
AGEH	0.025*	0.006			0.028*	0.008	0.029*	0.008	0.020*	0.006	0.021*	0.006
CITY	-0.401**	0.190			-0.007	0.055	0.031	0.059	-0.115**	0.046	-0.022	0.043
EDUC	-0.057	0.047			-0.080	0.048	-0.124**	0.045	-0.055	0.035	-0.045	0.037
FEDU	-0.048	0.040			0.001	0.060	-0.062	0.060	-0.035	0.046	0.020	0.047
MARR												
POOR			0.163*	0.067	0.181	0.155	0.112	0.151	0.161	0.155	0.078	0.121
RACE					-0.268	0.166	-0.028	0.169	0.078	0.130	-0.272*	0.129
SEXH					-0.178	0.153	-0.217	0.156	-0.119	0.119	-0.111	0.120
FOOD	0.015	0.019	0.12×10^{-3}	0.48×10^{-3}								
NCIG	0.064	0.045										
RELG	0.050	0.679										
RISK			0.028	0.025								
INSR			-0.012	0.087								
CHCH			-0.005	0.067								
COLD	0.001	0.002										
NOXT					0.0021*	0.0013						
NOXM	0.0047*	0.0023					0.0017	0.0012	0.0021*	0.0012		
NOXN											0.0013*	0.0007
SULT												
SULM	-0.0018	0.0032	0.0025*	0.0013								
SULN												
TSPT					-0.0008	0.0035						
TSPM	0.0078*	0.0038	0.85×10^{-3}	0.20×10^{-2}								
TSPN											-0.0009	0.0011
ULTV	-0.51×10^{-5}	0.16×10^{-3}										
Constant	-1.018		0.005		-0.378		-0.037		-0.285		0.005	
R^2	0.210		0.624		0.078		0.083		0.464		0.439	
S.E.	1.563		1.265		1.435		1.438		1.101		1.120	
F	(14,304) = 5.762		(9,340) = 62.58		(9,390) = 5.899		(9,390) = 5.899		(10,389) = 33.630		(10,389) = 30.490	
η_{NOX}	0.470		0.608		0.258		0.301		0.369		0.327	
η_{SUL}			0.514									
η_{TSP}	0.948											

*Significant at the 0.05 level of the one-tailed t-test.

**Significant at the 0.05 level of the two-tailed t-test.

[a] All observations in this table are limited to individuals who have always lived in one state, except for the observations in (2). These are limited to individuals who currently live within walking distance of relatives.

[b] The air pollution variables for this expression refer to arithmetic mean 1969-71 geometric mean concentrations in $\mu g/m^3$. The referenced 50 cities are 50 of the 60 cities used in the aggregate mortality study that form a part of this report.

coefficients to appear statistically non-significant when they are properly viewed as significant.

There are, however, two very important exceptions to the supposed absence of a multicollinearity problem: the types of air pollution tend to be very highly correlated and different moments of the same pollutant also are closely associated. As Table 5.5 shows, total suspended particulates and sulfur dioxide appear to have a very high linear association as do all the moments of a particular air pollutant. If one were to introduce nitrogen dioxide in Table 5.5, the linear association between this pollutant and total suspended particulates and/or sulfur dioxide would also be large, though somewhat smaller than that between the latter two pollutants. For example, in 1975, the simple correlation coefficient between various measures of total suspended particulates and nitrogen dioxide is never less than 0.50 and sometimes reaches intro the 0.70 or greater range. Given these close linear associations among the types of air pollution, we are reluctant to assign a health effect to a particular pollutant. Instead, it seems preferable to make the assignment to the outdoor air pollution phenomenon. In addition, when one or more air pollutants appear as explanatory variables in an estimated dose-response expression, the standard errors of each will tend to be somewhat inflated. Thus, a few of the air pollution coefficients to which we do not attach significance sometimes would be significant if one or more of the other air pollution variables were removed. Similarly, some of those air pollution coefficients that are significant would be more significant with the removal of a companion variable from the expression.

The above discussion does not deal with a dilemma posed by the issues of bias and multicollinearity. If the different types or moments of air pollution have separable impacts on health, then one biases the coefficients of the remaining explanatory variables by deleting one or more of the air pollution variables. Nevertheless, if one includes the highly collinear air pollution variables, he reduces the apparent statistical significance of any one of them. In this study, we do not directly attack the dilemma by constructing and then applying rigorous criteria for choice. We choose an easier and less rigorous course by estimating some expressions, each from a different sample, that include all the types of air pollution, while including only one type of air pollution in other expressions. To a very substantial extent, this course was forced upon us by circumstances: for some years over the nine-year SRC survey interval, there was no available information on particular types and moments of the air pollution variables.

Table 5.5 exhibits one other intercorrelation that is a cause for concern, namely a simple correlation coefficient of 0.70 between LDSA and DSAB, i.e., between the duration of a chronic illness and its self-reported severity. Relative to most other samples of the study, this intercorrelation is a bit low. For most samples, it is closer to or in excess of 0.80. Certainly, the length of a disease and its severity are not identical. In fact, one might expect those who are severely disabled to have relatively short disease durations: they are more likely to die. We may thus have increased the intercorrelation between these two variables by not making DSAB be monotonically increasing. The high intercorrelation

arouses suspicions about whether the two variables might be measuring the same thing, a clearly ridiculous state, if one is trying to explain the covariation between the two variables. Furthermore, if air pollution is expected to lengthen the duration of an illness, there is obvious reason to think that it will also make an illness more severe. More accurately perhaps, air pollution causes illness and increases the severity of preexisting illness, thus in a recursive fashion lengthening, for those who survive, illness duration. This implies that the estimated expressions which include DSAB as an explanatory variable are actually reduced form expressions, where DSAB is determined within the structural system. As a result, the single equation estimates with DSAB as an explanatory variable are not asymptotically efficient although they are consistent since DSAB is the only explanatory variable that would be determined within the structure of a recursive system. If instead of DSAB being a determinant of LDSA, it is actually another measure of the same thing in respondents' views, then DSAB must be dropped from the estimated expression. For the expressions estimated from some samples we include DSAB; for other samples, we delete it, using whichever of the preceding rationales conforms to the estimated expression. As we will see, inclusion or exclusion doesn't really make much difference to the signs and magnitudes of the coefficients for the major variables of interest, the air pollution variables.[11/]

In estimating dose-response expressions for chronic illness, we have used LDSA rather than (or in addition to) DSAB because only the former is stated in cardinal terms. LDSA, however, retains one disadvantage of DSAB; as presented on the SRC tapes, it takes on only five values. Although the first four of these values apply to approximate two-year intervals, the last value might better be termed "a long time," since it is meant to apply to disabilities lasting eight or more years. If one interprets, as we shall do in this chapter, this last value to be equal to exactly ten years, then the dependent variable for chronic illness has a measurement error that biases it downward, causing the effects of the explanatory variables to be underestimated. This could be a serious source of error since about 40 percent of those who are disabled in any given SRC survey year, or seven to eight percent of the total SRC respondent population, profess to have been disabled for eight or more years. Given this problem, which we disregard until a succeeding section, it is perhaps preferable to interpret the coefficients attached to the explanatory variables in the estimated chronic illness dose-response expressions as the proportion of one of the discrete values comprising LDSA associated with a one unit change in the relevant explanatory variable.

Yet another estimation issue is caused by the five discrete values assumed by LDSA, This small number of discrete values means that heteroskedasticity could be present in those expressions estimated by ordinary-least-squares techniques. Ideally, multinomial logit estimation would be employed; but because the number of parameters with multinomial logit estimation increases so dramatically when the dependent variable assumes more than two values, there is an explicit tradeoff between the misspecification possibly introduced by the use of ordinary-least-squares and the vastly increased cost and complexity of multinomial logit estima-

tion. We have opted here for simplicity and lesser cost while not dismissing the heteroskedadicity issue: we estimate the chronic illness dose-response functions by ordinary-least-squares hut peruse the estimated results. by simple graphic techniques to check for the presence of heteroskedasticity. Even if this undesirable property is present, it does not follow that our estimates will be biased and inconsistent. They will not be efficient (they will not have the smallest variance in a class of unbiased estimators), but they will be unbiased and consistent. The problem with heteroskedasdicity is thus not with the estimated coefficients themselves but rather with the calculated standard errors. These standard errors are biased, thus invalidating the tests of significance for the estimated coefficients.

There are a number of results for acute illness in Tables 5.6a and 5.6b worthy of explicit note:

1) Of the seven different unpartitioned samples used to estimate acute illness dose-response expressions, statistically significant air pollution coefficients occur in all of them. Thus, an additional unit of air pollution, as defined by any of the variety of measures employed here, was associated with an increase of from one to four hours in average annual hours of acute illness. Except for 1973, magnitudes of the air pollution coefficients are quite stable from one sample to another, even though the specifications for the expressions often differ substantially. No tests have been performed to establish whether there are statistically significant differences in the air pollution coefficients across samples.

2) For the unpartitioned samples, the elasticity, η, of acute illness with respect to any of the air pollution variables (a unitless measure of the response of acute illness to variations in air pollution) is substantially less than unity. This implies that in the immediate neighborhoods of the sample values of these variables, average annual hours of acute illness is increasing at a decreasing rate with respect to increases in air pollution.

3) Two of the four partitioned samples in Table 5.6b do not have statistically significant air pollution coefficients. If air pollution has any impact upon the frequency of acute illnesses among individuals who are chronically disabled and who live in families where a pack or more of cigarettes is smoked, the estimataion techniques and sample sizes employed here are incapable of capturing it.

4) When measures of total suspended particulates and sulfur dioxide are included as explanatory variables in the same expression, the coefficient for them usually assumes a negative sign. Generally, total suspended particulates will take on the negative sign. Similarly, when sulfur dioxide and nitrogen dioxide are included as explanatory variables in the same expression, nitrogen dioxide often assumes a negative sign. For estimated expressions in which total suspended particulates and/or nitrogen dioxide are used as explanatory variables, but which do not include sulfur dioxide, both of the former sir pollutants have positive signs. These sign switches could be due to the high linear associations among the pollution variables.

5) With some exceptions, an increase of one discrete value in either of the measures of chronic illness tends to increase the average annual hours of acute illness by from 20 to 40 hours.

6) With the sole exception of the variables for a poor childhood and race, the variables representing biological and social endowments fail to play a statistically significant and consistent role in the acute illness dose-response expressions. It is possible, of course, and perhaps even likely, that the race and childhood background variables are capturing many of the effects of low education, etc.

7) The life-style variables in the acute illness dose-response expressions consistently have the expected signs and are often statistically significant. This is particularly true for the exercise and nutritional adequacy variables: they reduce average annual hours of acute illness.

8) Contrary to expectations, the explanatory variable for availability of medical care, INSR, usually has a positive sign, implying that people with better access to medical care have more acute illness. We have no explanation for this other than a pure speculation that people with better access to medical care are more likely to recognize the symptoms of acute illness, perhaps because physicicians provide them with the information to recognize these symptoms. On the other hand, INSR might simply be a poor measure of the respondents' access to medical care.

9) Other than air pollution, only two alternative measures of the respondents' environments were employed as explanatory variables. These variables were used in only a limited number of samples. DENS, the number of persons per room in the respondents' residence, increased average annual hours of illness by more than three in the single sample where it was statistically significant. The variable for employment in the chemicals and metals manufacturing sectors had too small a number of individuals in each sample to yield statistically meaningful results.

10) Visual inspection of the residuals for expression (1A) of Table 5.6a and expression (1) of Table 5.6b did not reveal any serious heteroskedasticity problems.

We tentatively conclude from the preceding findings that the life-style and environmental variables, including air pollution, we have used probably play a significant role in acute illness. The evidence for the biological and social background and the access to medical care variables is substantially less clear both because of measurement problems and because racial differences in educational and childhood environment may be reflected in simple binary variables for race and a poor childhood. Finally, it should be noted that none of our expressions "explains" a very large portion of the variation in acute illness. The coefficients of determination never exceed 0.31 and are often about 0.10. Moreover, the constant term in each expression nearly always exceeds the sum of the coefficients of the explanatory variables. This is, of course, partly due to the scaling of the variables, but, given the number of binary variables (MARR, POOR, RACE SEXH, RELG, INSR), one might reasonably have not expected quite such a difference. The relatively unimportant role that many of the most statistically significant variables play in total variation in annual hours of acute illness is evident in the following partial coefficients of determination for variables appearing in various expressions of Table 5.6a: for expression (7), NOXM = 0.004, SEXH = 0.044, and and LDSA = 0.004; for expression (1B), SULM = 0.021, FOOD = 0.002, RACE = 0.043, NCIG = 0.029, and DSAB = 0.136; and for expression (5), TSPM = 0.013, POOR = 0.024, DSAB = 0.124. With no more than one or two exceptions, the two variables for chronic illness, LDSA and DSAB, made the largest contributions to explaining variations in annual hours of acute illness.

Tables 5.7a and 5.7b give the estimated dose-response expressions for chronic illness. The following features stand out in these expressions.

1. Of the twelve different partitioned and unpartitioned samples present in Tables 5.7a and 5.7b, air pollution coefficients are statistically significant in nine of them. Not all air pollution coefficients are statistically significant in the samples where more than a single air pollution variable appears, nor are the signs always positive for those air pollution coefficients that are statistically nonsignificant. No pattern similar to the negative signs that are attached to sulfur dioxide or other pollutants when sulfur dioxide is used as an explanatory variable in the acute illness dose-response expressions appears here, however. Of the samples having no air pollution coefficients statistically significant at the 0.05 level or better of the one-tailed t-test [expressions (1), (5), and (7) in Table 5.7a], all had air pollution coefficients with positive signs and t-values in excess of 1.0. Two of these samples [expressions (1) and (7)] had air pollution coefficients statistically significant at the 0.10 level of the one-tailed t-test. The magnitude (and signs) of the air pollution coefficients for expressions (1), (5), and (7) were similar to the magnitudes and signs of the air pollution coefficients for the other samples. They ranged between slightly less than 0.0020 and slightly more than 0.0045, with the bulk being between 0.0020 and 0.0030. This means that a change between 0.2 and 0.4 or 0.5 percent in one of the discrete values comprising LDSA is caused by a one-unit change in air pollution. In elasticity terms, these discrete values (index) of LDSA appear to be relatively unresponsive to changes in air pollution. Nearly all the elasticities of the discrete chronic illness index with respect to air pollution are in the 0.2 to 0.5

range, implying that a one percent change in air pollution generally causes a change in the index of substantially less than one-half of one percent. As was true for the acute illness dose-response expressions, this means that, in the immediate neighborhoods of the chronic illness index values and the air pollution values present in these samples, chronic illness duration increases at a decreasing rate with respect to increasing air pollution.

2. As earlier noted, translating the coefficients for the explanatory variables in the chronic illness dose-response expressions is invalid because the highest value in the index could, in real-time terms, be anything equal to or in excess of eight years. Nevertheless, if one assumes that the real-time involved in this last index value is equivalent to that in all the lower values, than the translation can be performed. With this assumption, the air pollution coefficients imply that an additional unit of air pollution is, on average, associated with an increase of from one and one-half to three and one-half days in the duration of chronic illness. As before, even with the aforementioned assumption, this rate is applicable only in the immediate neighborhoods of the chronic illness index values and the air pollution values present in the samples.

3. In those unpartitioned expressions where it is employed as an explanatory variable, the severity of the respondent's disabilities has a highly significant, positive, and strong effect on the duration of these disabilities. The partial coefficient of determination of DSAB with respect to LDSA was consistently about 0.40. The inclusion of DSAB in expressions did not appear to have an effect on either the magnitudes or the significances of the air pollution coefficients. Similarly, its presence or absence did not seem to make much difference to coefficients for the other explanatory variables.

4. Results for the biological and social endowment variables are mixed. Only respondent age is consistently significant with the expected sign. Generally, as expected, the level of the respondent's education is associated with lesser durations of chronic illness, but it is only occasionally significant. Poor parents tend to be consistently associated with increased chronic illness durations, but POOR is statistically significant in only one sample. Otherwise, variables such as CITY, FEDU, MARR, RACE, and SEXH contributed very little to the expressions. Rarely were they significant statistically. More importantly, their magnitudes and their signs proved to be expremely sensitive to whatever specification was adopted.

5. Because it is not clear that the magnitudes of lifestyle variables are independent of the duration of chronic illness, fewer of them were used, and those that were used were used less frequently, than in the acute illness dose-response expressions. EXER is an obvious case and it has not entered the chronic illness expressions. In fact, except for RELG, food adequacy is the only explanatory variable that enters the expressions estimated for more than one sample. It always has the expected sign but is never quite statistically significant at the 0.05 level selected for this study. On the rare occasions when they appear, both cigarette consumption and fundamentalist religious affiliations have the expected signs. RELG in expressions (7A) and (7B) just barely misses being crowned with statistical

respectability. Since religious affiliations seem likely to remain unchanged whether or not one is disabled, this variable probably should have been included for the expressions estimated from each sample.

6. The variable representing the availability of medical care, INSR, performed well for those four samples where it was used. Its sign was consistent with an interpretation that medical care availability reduces the duration of chronic illness. Unfortunately, its sign is also consistent with another interpretation: those who are chronically ill have difficulty procuring medical insurance. This latter interpretation means that INSR could be a function of LDSA. Given these conflicting interpretations, and having no information on which interpretation is likely to dominate, we have compromised and included INSR in some expressions while neglecting it in others. Its inclusion or exclusion does not appear to have any discernable effects on the coefficients for the air pollution variables.

7. Of the environmental variables, only CHEM seems worthy of comment. In the one expression where they appear, neither COLD nor ULTV were statistically signficant although COLD did have a positive sign. The statistical significance of CHEM in expression (3) of Table 5.7a should be disregarded. Expression (3) was estimated from a sample having only three people employed in the chemicals and metals manufacturing sector. None of these three people had a chronic disability.

8. With the exception of DSAB, none of the included explanatory variables explain substantial proportions of the variation in the index for duration of chronic illness. The air pollution variables, taken together, explain no more than two percent of the variation in LDSA; AGEH sometimes explains as much as five percent and EDUC usually explains around three percent of this variation. As with the acute illness dose-response functions, we have not been able to account for very much of the sources of variation in chronic illness.

9. Table 5.7b exhibits the estimated expressions for samples that were restricted to the values of the variables indicated at the top of each column. Contrary to similar restrictions placed on the samples for the acute illness expressions, these restrictions did not alter the explanatory variable coefficients in any noticeable fashion.

10. The patterns of the residuals for several of the expressions in Table 5.7a have been visually inspected for evidence of heteroskedasticity. When this Problem is present, it appears that the residuals tend to increase with increasing values of the dependent variable. Because the highest discrete value of LDSA has no upper bound, it is likely that the true variance of the sample tends to increase with increasing values of LDSA. As Kmenta (1971, p. 256) shows for expressions with a single explanatory variable, if the residuals and the sample variance are positively associated, the standard errors of the coefficients for the explanatory variable will be biased downward, causing the t-value to be too great. This need not be true, however, for expressions with multiple explanatory variables. The extent to which this has resulted in exaggeration or underestimates of the levels of significance for the chronic illness dose-response expressions is presently

Table 5.8

Lagged Effects of Total Suspended Particulates upon Duration of Chronic Illnesses (LDSA) of Respondents Who, as of 1975, Had Always Lived in the Same State

	(1) Unweighted		(2) Weighted	
	β	s	β	s
AGEH	0.012*	0.004	0.017*	0.005
EDUC	-0.009	0.040	-0.103*	0.050
MARR	-0.331*	0.160	-0.237	0.232
POOR	0.150*	0.110	0.327*	0.153
SEXH	-0.012	0.023	0.046	0.235
FOOD	-0.035*	0.021	-0.076	0.074
RELG	-0.003	0.030	-0.501*	0.286
CHEM	0.249	0.247	-0.147	0.332
TSPM5	0.4×10^{-4}	0.061	0.002	0.003
TSPM4	0.001	0.033	0.001	0.005
TSPM3	0.001	0.003	-0.001	0.004
TSPM2	0.003	0.016	-0.003	0.004
TSPM1	0.008*	0.004	-0.003	0.005
TSPM0	0.007	0.011	0.004	0.005
TSPM9	0.006	0.006	0.002	0.005
Constant	0.444		-0.690	
R^2	0.184		0.129	
S.E.	1.032		0.969	
F	(12,347) = 6.481		(15,210) = 4.082	

*Statistically significant at the 0.05 level of the one-tailed t-test.

unknown. The heteroskedasticity appears to be by far the most prominent for those estimated expressions having coefficients of determination less than 0.10.

It is widely thought that pollution-induced chronic illness is usually the result of cumulative, rather than instantaneous, exposures. Previously we have taken the position that, if only non-movers are represented in the sample, air pollution exposures during the year for which the respondent reports his behavior and status serve as adequate proxies for differences among respondents in cumulative exposures. If this position is at all tenuous, we have available the data to remedy it at least partially; that is, we have available information on respondent residential locations and air pollution exposures for a number of years. Table 8 presents some preliminary results involving an attempt to estimate the lagged effects of total suspended particulates upon the duration of chronic illness for 1975 respondents who have always lived in the same state. Since it is unclear exactly what a lagged effect of pollution upon the duration of an illness means, we exploit the high intercorrelation between LDSA and DSAB and interpret the expressions in terms of the lagged effects of air pollution upon the severity of chronic illness. As in earlier tables, the integers attached to the acronym for mean total suspended particulates refer to the year. Thus, for example, TSPMO refers to particulate concentrations in 1970.

The expressions presented in Table 8 have involved no tinkering: these are the first expressions having LDSA as a dependent variable that have used either of these samples. Expression (1) is an unweighted lag in which earlier air pollution concentrations are simply entered as additional explanatory variables. In spite of the very high simple correlation coefficients ($\simeq$ 0.80) among the air pollution values of the various years, at least one year (1971) is statistically significant. Moreover, the magnitude of the coefficient increases from 1975 to 1971, and then starts to decline. We have no explanation for this rather neat pattern and tend to suspect that its very neatness in an anamoly that would fail to emerge in expressions estimated from other samples. These other samples have not yet been exploited.

The air pollution series in expression (2) has more structure imposed on it. In particular, the series is assumed to follow a geometric lag distribution where the coefficients decline in fixed proportions, causing the impact of more distant air pollution concentrations to become progressively smaller. Clearly, expression (2) does not accord any importance to total suspended particulates. However, this does not mean that all weighted lag structures will give similar results. Estimation techniques are available that allow one to fit polynomial structures of any degree. These techniques have not been applied here.

In concluding these remarks about dose-response functions, we must make explicit a feature of the SRC data set that could readily cause the morbidity effects of air pollution and other negative health influences to be biased downward. This possible bias is due to the retrospective feature of the SRC data: living individuals are questioned about their behavior and status during the preceding year. The problem arises because

some potential respondents who were alive during the preceding year are dead by the time the interview occurs. Presumedly, those who died would tend to be those who were most seriously ill. If air pollution and other negative health influences contribute to this seriousness, or if those who are most seriously ill are most sensitive to air pollution, then the health impact of air pollution will be understated. Thus, the dose-response functions presented here are relevant only for those individuals who managed to survive over the time interval which the interviews described and the calendar date at which the interviews occurred. This qualification applies to all sections of this report where the SRC data is exploited. It is not a minor qualification since approximately five percent of the respondents died between interview years.

5.4 Recursive Estimates of the Effect of Air Pollution Upon Health, Labor Earnings, and Hours of Work

In the past decade, a number of empirical studies have appeared that describe the effect of health status upon labor productivity, where productivity effects are measured in lost earnings and work-time.[12/] At the same time, numerous epidemiological studies that attempt to associate health status with air pollution have been published.[13/] Thus far, no one has tried to combine the two study objectives in order to grasp the effect of air pollution upon either of the aforementioned measures of labor productivity. This section is a first attempt to do so. Labor productivity effects have never been explicitly included in quantifications of the benefits of national air pollution control efforts. Our results suggest that these productivity effects could constitute a significant portion of these benefits and are certainly worthy of further study.

In spite of a number of limitations which will later be exposed, the section has at least three unusual, if not utterly novel, features. First, although it treats health status as an exogenous rather than endogenous variable, a structural equation for health status is specified. This contrasts with nearly all epdemiological studies, where the analysis is confined to reduced-form health status, making any direct assignment of health effects to air pollution an extremely tenuous operation. Second, the health parameters in this section are estimated in the context of structural expressions for hourly earnings and annual hours of work. Finally, possible differences in effects of air pollution upon crude measures of acute and chronic generalized health status are recognized. The null hypothesis is that air pollution, by enhancing initial susceptibility and by making recovery more difficult, causes acute and/or chronic health problems. This, of course, was the theme of the previous section. In this section, we wish to ascertain the impact, if any, of these air pollution-induced health adversities upon earnings rates and hours worked. Thus through the intermediary of any health problems it induces, air pollution can be said to indluence labor productivity.

Even though health is treated as being exogenously determined, the Grossman (1972) model of health production can serve as the analytical foundation of the expressions to be estimated.[14/] This model views the individual as the producer, via his selections of mixes of market-purchased goods and his own time, of health status. Within the context of this approach, earnings

Table 5.9

Simple Correlation Coefficients Between Labor Supply and Certain Other Variables for a 1970 Sample

	WAGE	WORK	BDALO	LTWK	ICTR	UION	RINC
WORK	0.085	1.000	0.287	0.123	-0.629	0.101	0.479
BDALO	0.235		1.000		-0.167		0.656
LTWK	-0.038		0.070		0.155		0.054
UION	-0.039		0.108		-0.131		0.070
RINC	0.465		0.656		-0.268		1.000
ICTR	-0.421	-0.629	-0.167		1.000		-0.268
ACUT	-0.042	0.012	-0.170		-0.102		-0.079
DSAB	-0.134	-0.468	-0.156		0.325		-0.227
LDSA	-0.141	-0.441	-0.143		0.303		-0.190
AGEH	0.017	-0.174	0.197		0.156		0.107
CITY	0.094	0.016	0.018		-0.057		0.092
EDUC	0.165	0.323	0.493		-0.172		0.465
FEDU	0.044	0.131	0.244		-0.132		0.148
FMSZ	-0.038	-0.046	-0.424		0.165		0.094
POOR	-0.114	-0.116	-0.126		0.094		-0.058
RACE	-0.014	0.008	0.139		-0.199		0.018
SEXH	0.072	0.448	0.360		-0.284		0.480
EXER	0.072	0.161	0.193		-0.150		0.217
FOOD	-0.007	-0.078	-0.058		-0.239		-0.061
RISK	0.139	0.235	0.538		-0.138		0.427
INSR	0.088	0.505	0.439		-0.413		0.440
CHEM	-0.012	-0.033	-0.082		0.040		-0.022
SULT	-0.037	-0.174	-0.163		0.085		-0.183
SULM	-0.038	-0.137	-0.113		0.083		-0.134
SULN	-0.056	-0.109	-0.077		0.048		-0.127
TSPT	0.002	0.081	0.087		0.066		0.075
TSPM	0.046	-0.005	0.086		0.044		0.130
TSPN	0.058	0.009	0.122		0.046		0.170

rates depend on various forms of investment in human "capital" (e.g., education, prior lifestyles, and medical inputs) and labor market conditions; and the time supplied to the labor market depends on the individual's hourly earnings and the quantities of goods and time desired for household production and consumption. Health states depend on the prior resources the individual has devoted to their production.

Except for certain of the environmental variables, the data used? to estimate the model consist of four distinct samples drawn from the 1969, 1970 and 1971 SRC interview data. Several variables, defined in Table 5.1, are used in this section that were not used in the preceding section. For one of the samples, Table 5.9 provides the simple correlation coefficients between these additional variables and some of the other previously used variables. Representative means and standard deviations for the additional variables are available in Table 5.2.

Table 5.9 gives little attention to LTWK and UION because our major interest in them is their association with WORK, WAGE, and RINC. Absenteeism was checked in this sample but apparently none of the respondents would admit to being absent from work for reasons other than sickness. As was noted in Table 5.4, where 81.1 percent of the respondents had annual asset incomes of no more than $500, most of the respondents' annual incomes not earned during the current year appear to be governmental transfer payments. This accounts for the negative and high correlations between ICTR and RINC and WAGE. Note also in Table 5.9 that the simple correlations between the two chronic illness measures, DSAB and LDSA, and WORK and RINC are quite high.

The household head's annual number of work hours, WORK, and his hourly earnings, WAGE, are used as the empirical representations of the endogenous variables in the model. Remember from the definitions of Table 5.1 that WAGE is an approximation of the marginal, rather than the average, wage rate. Annual number of work hours is used as the sole measure of labor supply because the sample contains no information on the seasonal distribution of hours when working. Neither vacation time nor sick time is included in annual hours worked, even if the individual was paid during these times.

The system to be estimated for each sample consists of four expressions: a chronic illness expression; an acute illness expression; a wage expression; and a labor supply expression. A representation, in implicit form, of this structural system is as follows:

1. LDSA = f(Biologica1 and social endowments, Lifestyles, Medical care, Environmental). (5.1)

2. ACUT = g(LDSA, Biological and social endowments, Lifestyles, Medical care, Environmental). (5.2)

3. WAGE = h(LDSA, ACUT, Cost-of Living, Experience, Biological and social endowments). (5.3)

4. WORK = k(WAGE, LDSA, ACUT, Transfer income, Wealth). (5.4)

As structured, this system is obviously recursive.

A great deal of research is available [e.g., Lazear (1976)] showing that earnings are positively related to formal and informal schooling. Good health is here viewed as having effects on earnings analogous to the effects of increased schooling; that is, good health increases the individual's marginal value productivity and therefore raises his marginal earnings. In addition, previous good health may have had an indirect effect on earnings by easing the task of achieving schooling success and thereby ultimately improving the individual's productivity and associated earnings. The EDUC and LOCC variables in (5.3) are intended to capture the effects of training upon earnings. They may also reflect, in part, the influence of past health status. The health status variables, ACUT, DSAB, and LDSA, in (5.3) register the effect of current health status, via the effect on productivity, upon earnings. Since chronic illnesses reflect long duration, as opposed to temporary, reductions in productivity, we expect wages to exhibit greater responsiveness to the chronic illness variables than to the acute illness variable.

In addition to the aforementioned variables, the marginal earnings expression includes variables representing the 1970 cost-of living in the county of residence as well as variables representing the individual's race and sex. If, as is frequently asserted, being non-white or female negatively influences marginal earnings, either labor market discrimination or less market productivity in the current period could account for the influence.[15/] The structural system we employ is incapable of distinguishing between the two possible influences.

Cost-of-living, BDAL, in the county of residence is accounted for in (5.3) because it is real marginal earnings, rather than money earnings, that limit the extent to which the individual is able to satisfy his cravings and yearnings.

As Mincer (1970) and others have shown, earnings expressions similar to (5.3) should be semi-logarithmic, where the dependent variable is the logarithm of the earnings term. In this paper, we presume the earnings expression to be linear in the original variables. This presumption was adopted in order to obtain a sample of individuals possessing reasonable variability in the values of the health variables, earnings, and hours worked. If, in order to avoid having to assign positive earnings to individuals who really had zero earnings, only individuals who actually had positive earnings were included in the sample, the variability of the chronic disability measures would have been substantially reduced, thus requiring that inferences about the influence of air pollution on health, earnings, and hours worked be drawn from the relatively few remaining individuals whose health status and work patterns differed substantially from the mean. Moreover, dropping individuals with zero earnings from the sample would have meant that those individuals with long-standing and/or severe chronic health problems would be excluded.

Expression (5.4) the annual hours worked or labor supply expression, is consistent with the treatments of health capital in Grossman (1972). Improvements in health states increase the total time available for work and for consumption. With real earnings and consumption opportunities held constant, the consumer would be inefficient, assuming he was initially in equilibrium, if he allocated all this additional time solely to consumption. This is because the ratio of consumption time to work would rise, causing the marginal value of consumption time to become less than the marginal earnings that could be obtained. To recover equilibrium the individual would have to devote the additional time to both work and consumption. We therefore expect the amount of work time to increase with improvements in health status.

In addition, since health status is assumed to be exogenous, an improvement in health would increase the wage rate as well as the pecuniary equivalent of time spent in consumption. In terms of the household production approach to consumer theory, "full income" would be increased. The health improvement therefore would constitute a pure income effect, causing the individual to increase the value he attaches to any unit of consumption time. This increase in the value of consumption time would cause the individual to increase his demand for those marketed goods permitting him to use this more highly valued consumption time with greater effectiveness. The purchase of these marketed goods requires that he obtain more income, and therefore that he increase his work time.

An increase in income not earned in the current period, ICTR, would also result in a pure income effect. However, because the additional income is not a consequence of improvements in work productivity, the value of work time relative to consumption time decreases, assuming the wage rate and health status to be invariant. The result is that with an increase in income not earned in the current period, the individual must reduce work time in order to restore equilibrium.

The preceding remarks indicate why the sign of the marginal hourly earnings variable, WAGE, in (5.4) is ambiguous. An increase in marginal hourly earnings would increase the value of work time relative to the value of consumption time, causing the former type of time to be substituted for the latter. However, the increase in marginal hourly earnings has simultaneously increased the individual's "full income," causing the value he attaches to any given unit of consumption time to increase. Whether the increase in the value of consumption time exceeds the increase in the value of work time is an empirical question.

Since the immediately preceding remarks refer only to real marginal hourly earnings, (5.4) includes BDALO, the cost-of living index, in order to control differences in real earnings among counties of residence.

The four-equation system, in which acute and chronic illnesses are exogenously determined, represents a strictly recursive system. First, health status is determined independently of hourly earnings and hours worked, and then health status is used to determine hourly earnings and hours worked. Similarly, hourly earnings are determined independently of